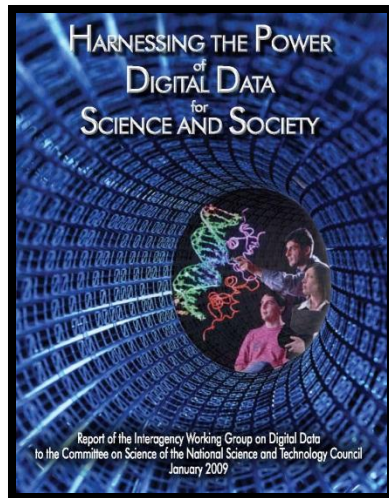# HARNESSING THE POWER OF DIGITAL DATA: TAKING THE NEXT STEP

Report of the Scientific Data Management Workshop

Co-sponsored by EPA, CENDI and IWGDD

June 29-July 1, 2010

Workshop Objectives

IWGDD Harnessing Report

- Address two key objectives of the IWGDD Harnessing Report
  - Agencies should have a data policy
  - Projects should have data management plans
- Bring in the Practitioner perspective
  - Policy analysts, operational user, researcher, science manager + data manager
  - Share experience and best practices
  - 87 people from 25 agencies attended

# Workshop Structure Set to Stimulate Discussion

- Opening Plenary
  - To set the policy stage
  - To provide common background from survey results
- Technical Plenary
  - Best practices of 6 agencies covering diversity of data types
- General Discussion
- Breakout Sessions
  - Policy
  - Plans
- Closing Plenary
  - Reactions to Breakout Reports and General Discussion

EPA Chief Scientist
Dr. Pai-Yei Whung

Substantive Background Planning

- Planning Team from Sponsoring Agencies
- Regular discussions
- Work products
  - Agenda
  - Background reading list of material on data management policies and plans
    - 3 Key Documents
  - Extensive Baseline Survey of knowledge and attitudes
  - Matrices of Elements of Policies and Plans

Does your agency manage data as an enterprise asset? (Q8, N39)

14, 36%

25, 64%

Yes No
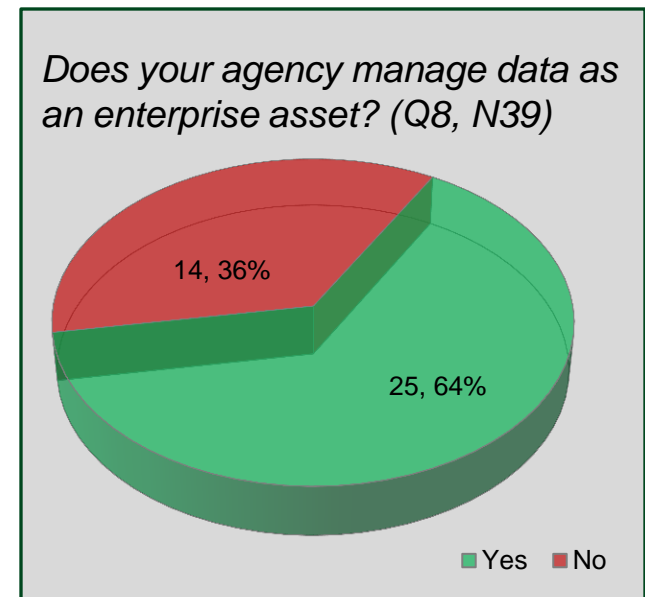
Report Preparation

- Initial Input from Volunteers

- Drafting Team developed initial ideas

- Editors pulled together first draft

- Ready for sponsor review (January)

- Attendee review of the full draft after sponsor input (February)
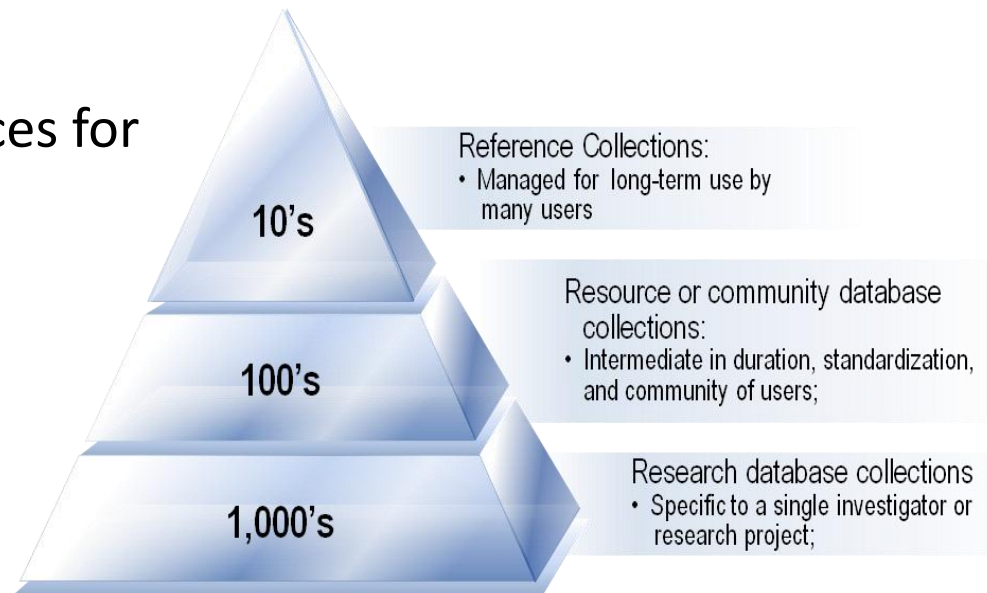
- Final report (March 2011)

Report Outline

- 1.0   Introduction

- 2.0   Data Policy

- 3.0   Data Management Plans

- 4.0   Recommendations

- 5.0   Conclusions and Next Steps

    - APPENDIX C: SAMPLE Data Policy for Digital Science Data Based on Federal and Industry Best Practices

    - APPENDIX D: PROJECT Data Management Plan Template

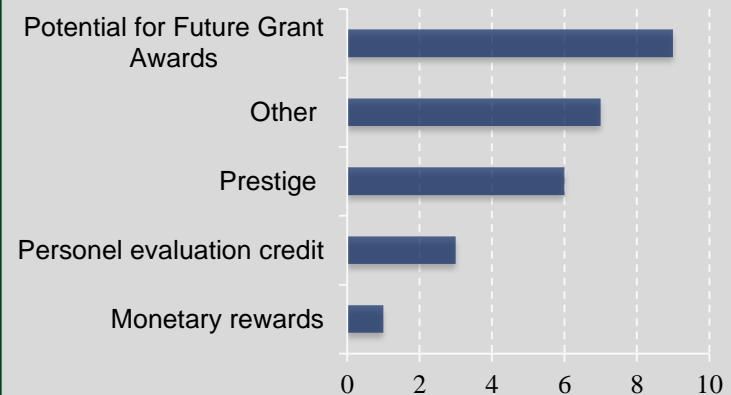# Recommendations to Federal Policy Organizations (OSTP, OMB, OPM) - 1

- Need for better understanding of value proposition
  - To better manage the data asset

- Portfolio Management as a Tool
  - To allocate resources for management and preservation
  - ROI goals and risk management

Reference Collections:
- Managed for long-term use by many users

10's

Resource or community database collections:
- Intermediate in duration, standardization, and community of users;

100's

Research database collections
- Specific to a single investigator or research project;

1,000's

Recommendations to Federal Policy Organizations (OSTP, OMB, OPM) - 2

- Stimulate Culture Change through Incentives to Stakeholders
  - Changing concept of ownership
  - Needs some top down leadership
- OPM/Agency Reward Structure should Reflect Data Policy Objectives
- OPM Should Revise Standards for Researchers
  - Establish Data Management Series
- Training Should be Developed to Leverage Personnel Resources

In your community are there incentives that promote data sharing? (Q33, n=36)

Other: In-kind sharing, Promoting data sharing relationships (2x), Senior management directive, Incentive to develop a better product (2x)

Recommendations to Federal Policy Organizations (OSTP, OMB, OPM) - 3

- Interagency Coordination Function Should be Established
  - Built on IWGDD
  - Must Connect to other sectors and international activities
- A Scientific Research Agenda should Provide better objective foundations for SDM Decisions
- Reflect Open Government Goals in Data Management Policy and Planning
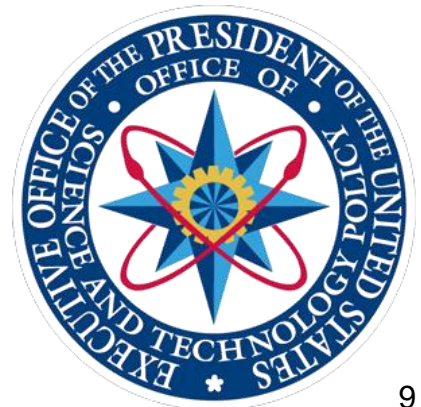
Recommendations to Agencies on Data Policy- 1

- Each Agency should have a Data Policy that is Harmonized with other Agencies and Compatible with Programmatic and Community of Practice Policies.
  - Agency data policies should reflect federal-wide Guiding Principles (Harnessing Report)
  - Agency data policy should ensure that Scientific Data Management Processes are Integrated With Knowledge Management Initiatives
  - Agency Policy and Structure of Responsibility Should be Clear Regarding Data Retention

# Recommendations to Agencies on Data Policy- 2

- Agencies Should Manage Scientific Data for Appropriate Control but make Appropriate Access More Transparent.
  - Landscape of restrictive categories difficult to navigate
  - Dearth of long-term repositories providing access under a homogeneous classification system
- Agencies Should Establish the Role/Function of Chief Data Officer
  - Clarify Roles and Responsibilities of Agency Personnel to ensure stewardship
  - Establish relationship with CIO, Enterprise Architecture, etc.

In your community of practice, do you have controls for unclassified but restricted information? (Q37, n=35)

6, 17%

22, 61%

8, 22%

- Sometimes
- No
- Yes
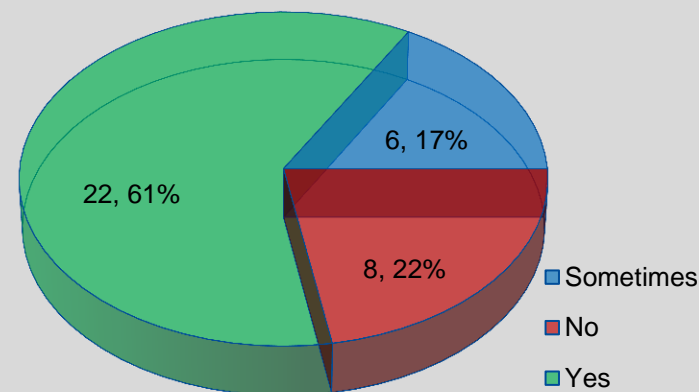
Recommendations to Agencies on Data Management Planning

- Scientific data should be managed according to an SDM plan that covers the full Data Life Cycle and integrates with the Project Life Cycle

  - Project Life Cycle – Sustainability and context

  - Data Life Cycle – Long-term preservation/transfer of responsibility

- The Data Management Plan should be "an ongoing open ended living electronic Record" that follows the data through its life cycle.

  - Exploration of best DMP implementation approaches

*IWGDD's digital data life cycle model*

(NSTC, 2009)
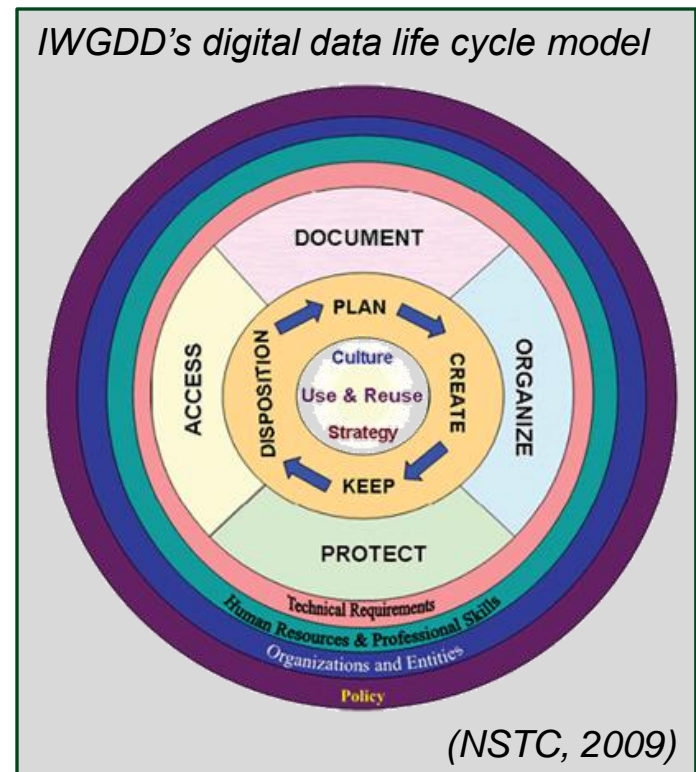
Recommendations on Communities of Practice

- Agencies Should Make Effective Use of Communities of Practice
  - Interagency workgroup should develop a Community of Practice Conceptual Framework for Scientific Data management
    - What it is, how to implement it, unique aspects of science data management
    - Explore data management cost sharing models
  - Agencies Should Support The Role Of Communities Of Practice
    - Increased Outreach to help Primary Data Generators Identify Future Users Of Their Data.

Recommendations on Infrastructure

What (3) elements of a SDM Plan would be most helpful to you in evaluating data for use in making and defending policy recommendations. Please only indicate three. (Q29, n=36)
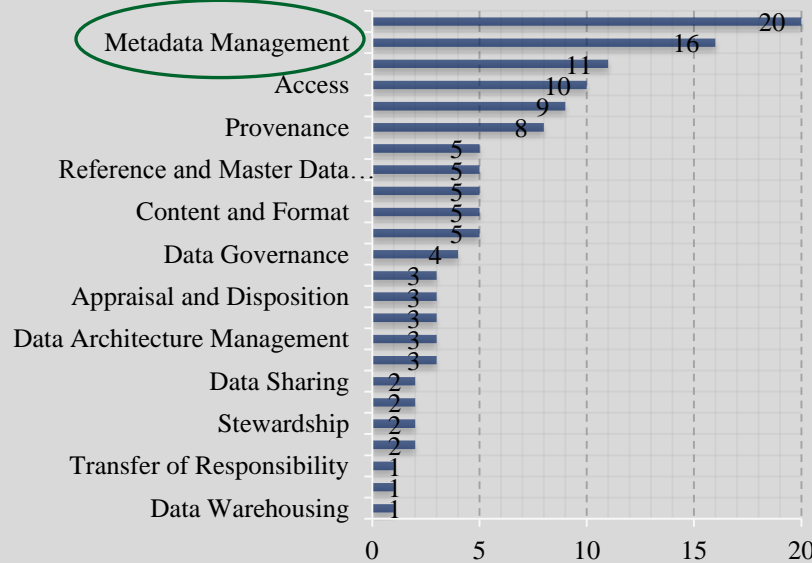
- Agency Infrastructure Should Support Scientific Data Management
- Policy and Individual Agencies Should Support Persistent Identification Across the Government
- Includes Version Control to Facilitate Data Management and Use
- Data Management Planning In and Across Agencies Should Include a Commitment to Effective Metadata Management

Conclusions and Next Steps

- Growing recognition of the importance of data management
  - From Scientific Integrity to PCAST report on the Digital Future
- Continuing the dialogs is critical
- Sponsors each have a report application
  - CENDI to promote work within agencies
  - IWGDD to continue interagency coordination
  - EPA to apply to developing best practices

Questions?

Your input is welcomed